Progress Report

Period Covered by the Report
• *10 November 2001 to 10 December 2001*

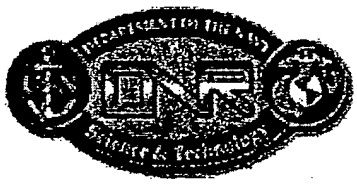
Date of Report
• *10 December 2001*

Project
• ***Compadre: A Device Independent Voice-to-Voice Language Translator Software Solution***
• SBIR Phase I Topic N01-044
• Contract Number N00014-01-M-0225

Item Number
0001AD: Progress Report

Security Classification
Unclassified

Sponsor
Dr. Joel Davis
Office of Naval Research
Ballston Tower One
800 North Quincy Street
Arlington, VA 22217-5660



Principle Investigator:
Robert Palmquist
SpeechGear, Inc
3454 Circle Bluff Court
Faribault, MN 55021

Program Partners

VISUAL GOLD

IRISUSA

ARAMEDIA

L&H/Apptek

CiMOS

Compadre:
A Device Independent Voice-to-Voice Language Translator Software Solution

---

A Phase I SBIR Speech Application Project
for the Office of Naval Research

SBIR Call Number N01-044
Contract Number N00014-01-M-0225

SPEECH GEAR

Robert Palmquist
SpeechGear, Inc
3454 Circle Bluff Court
Faribault, MN 55021

Telephone: 507-664-9123
Email: rpalmquist@speechgear.com
Web: www.speechgear.com

A. Project Summary

Technical Abstract:

Mission Statement
To develop and deploy language translation software that is device independent, supports bi-directional translation of multiple languages, produces text transcriptions of spoken conversations and supports translation of text extracted from digital images. This software shall run in both a reduced functionality standalone mode, and by wirelessly connecting to remote servers, a full-function mode. This software shall run on multiple pocketable platforms resulting in a mobile system that is low in cost, easy to use, robust in operation and comfortable to carry and/or wear.

The object of this Phase I research effort is to investigate the scientific, technical and commercial merit and feasibility of the system described in the preceding mission statement. Specifically, the team will investigate design options for the mobile translator system, identify potential applications, and select the best option(s) to pursue in making the design a reality. Four technical areas will be investigated: potential pocketable computing platforms, the operator interface, optical character recognition software and the language translation software. The commercial feasibility of this design will also be investigated. This includes identifying potential applications, languages to be supported, cost, and user requirements such as interface modes and response times. By combining both the commercial and technical elements, a complete definition of successful software and system solutions for pocketable language translation devices will be achieved.

Prototype systems showing device independence will be developed and demonstrated and a final report written documenting the Phase I results and recommendations for follow-on research and development in Phase II. Options are included for incorporating additional language pairs into the system and application specific terminology.

Anticipated Benefits/Potential Commercial Applications of the Research or Development:

Applications include all individuals who require multi-lingual capabilities. The mobile translator will benefit a wide range of individuals including military personnel, airport employees, border patrol and customs agents, police, fire fighters, retail clerks, bank tellers, delivery personnel, phone operators, tourists and any industry that sells, develops or manufactures products to/in global markets or employs individuals that do not speak the native language.

B. Project Status

B.1 Status Overview:

The overall work breakdown structure is provided in Figure 1. For purposes of this report, the project start date is selected at The actual purchase order was not received in the mail, however a FAX copy of the signed document was provided by Jennifer Schoen on

As is shown in Figure 1, a successful demonstration of the English/Arabic proof-of-concept system was given at the Office of Naval Research on November 26, 2001. This included all three usage modes: standalone, camera-based and voice-based. The demonstrations were performed commensurate with the Design Requirements (DR) and Prototype System Design (PSD) documents that were developed during the course of this Phase I effort with the only exception being that the voice-based system was demonstrated using a laptop versus using telephones to connect to a remote server. The DR, which is included in Appendix A of this report, contains the targeted and desired specifications for *Compadre's* overall system performance. This document was submitted in the July progress report and was approved per telephone conversations with Dr. Joel Davis. The PSD document, which is included in Appendix B of this report, contains a description of the overall system design. This document was submitted in the September progress report and was subsequently approved. In short, the DR describes what the system does, whereas the PSD describes how this is accomplished. The one critical item that remains is to use a telephone to collect spoken phrases versus a microphone headset. The required hardware (e.g., TAPI modem) has been evaluated, procured and installed. The software components have also been either acquired or written. Work is continuing to achieve this capability with a targeted completion date of December 24, 2001.

B.2.2 Camera-Based Mode

There are situations where using a touchscreen or keyboard to input foreign text will not be practical. One such example is the sign containing Arabic text that is shown in Figure 8. In this situation, it would be very difficult for an English-only speaking individual to enter the Arabic text using a keyboard or touchscreen or to look-up this text in a traditional English/Arabic dictionary. The same situation is present for multiple languages such as Korean, Japanese and Russian. To help solve this problem, *Compadre* allows the user to input text into the SmartPhone using a digital camera. A patent application for this capability has been submitted. The design of the prototype system is shown in Figure 9. Two different cameras are being used: a compact camera from HP that is very convenient to use and a high resolution camera from Minolta with superior capabilities but a more involved interface. The Minolta *Dimage 7* is being used to develop translation capabilities for full text documents with small font sizes (e.g., a complete page of Arabic text) whereas the HP camera is used for larger font sizes such as signs.

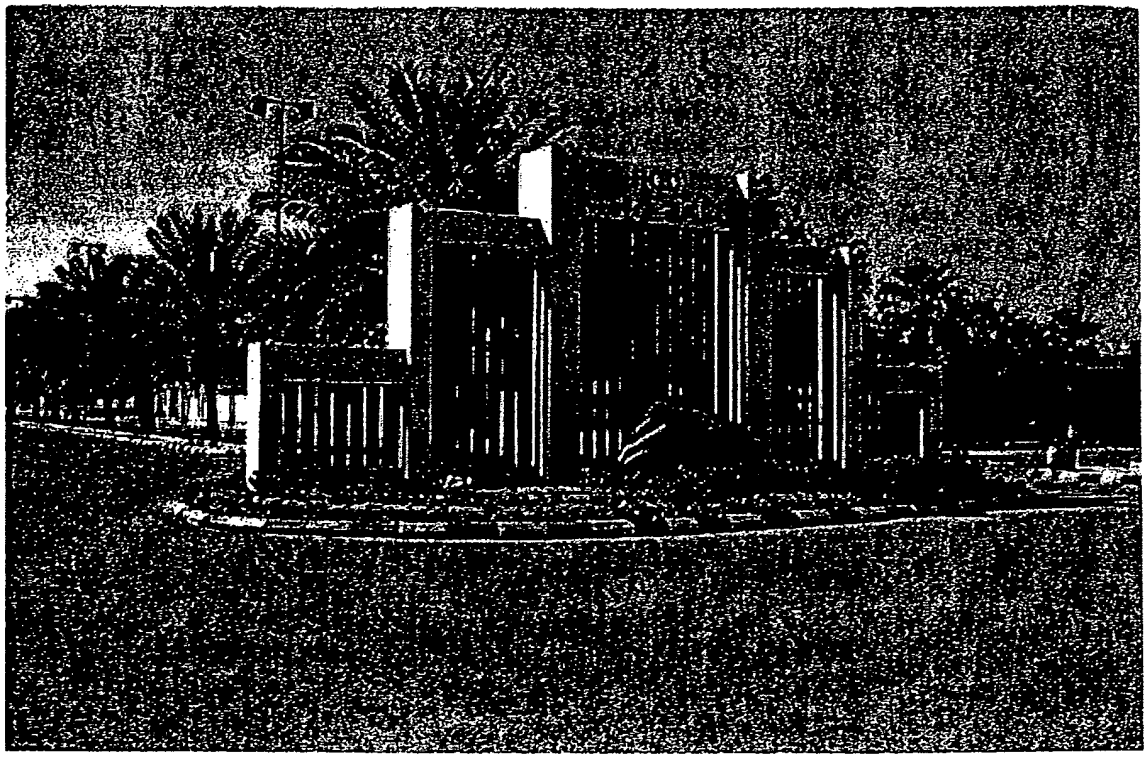


Figure 8: Examples of Arabic Sign

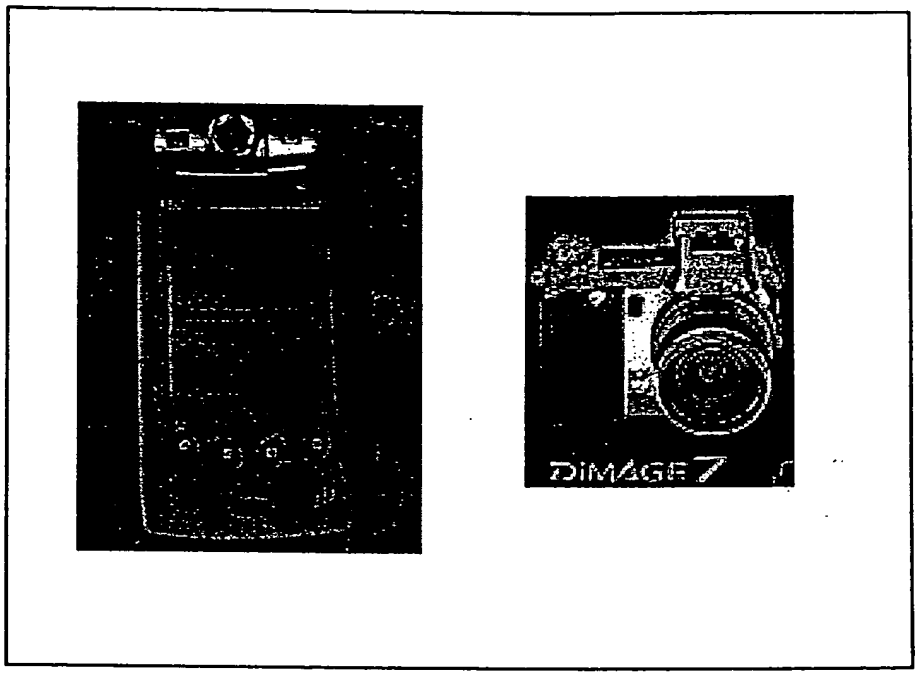


Figure 9: Examples of Camera-based Systems

Note that *Compadre's* software is designed to be device independent, thus, these are just two of many hardware configurations that could be used for this usage mode. One interesting alternative device is Samsung's conceptual product of including a camera with a cellular phone. This product is shown in Figure 10.

The digital camera is used to capture an image of the foreign language as is shown in Figure 11. Once the desired image is obtained, a "one-click" GUI is used to wirelessly connect the SmartPhone to a remote server where the image will be processed and the resulting translation sent back to the user. This is shown in Figure 12. This process takes approximately one minute to complete with the vast majority of this time being consumed by uploading the image to the server. Status bars, which are shown in Figure 13, are displayed to inform the user as to the percentage completion of each of the uploading and downloading procedures. The resulting translation is then provided along with the original picture. An example of this is shown in Figure 14. Note that for most situations the wireless connection will be made using cellular telephones. Because of the limited bandwidth of such a connection, it is important to reduce the overall size of the transmission. Thus, SpeechGear evaluated different image compression algorithms and selected the *Imagist* product from Visual Gold. SpeechGear is currently embedding *Imagist* directly into SpeechGear's software. This, along with several other features SpeechGear will implement in Phase II, will significantly reduce the time it takes to

upload images and thus reduce the overall time it takes to complete the translation process.

An additional user screen is accessed by selection the "tools" tab, which is located at the bottom of the user interface (see Figure 14). This screen, which is shown in Figure 15, is used to specify parameters, such as the host address, user account and password, of the remote server that *Compadre* is using to perform the translation process. Individuals can use this tool in the field to establish connectivity with additional severs. For example, if the a laptop is residing in a vehicle, or a soldier's has a wearable computer, the user could redirect the connectivity to this nearby platform and use Infrared or 802.11 to provide the connectivity versus a cellular telephone.

For the Phase I proof-of-concept system, the following phrases, in Arabic, have been included in the system:
"Hospital"
"Speed Limit 50"
"No Parking"
"Grocery Store"
"Post Office"
"Telephone"
"Emergency Use Only"
"Authorized Personnel Only"
"Danger, Do Not Enter"
This set of possible signs was selected to place a boundary on the overall scope of the OCR software requirements. In Phase II this limitation of preselected phrases will be removed.

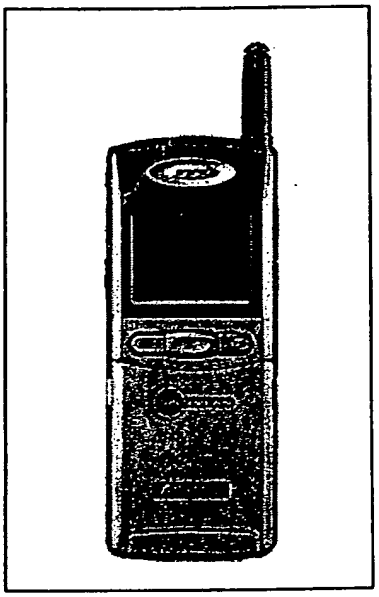


Figure 10: Samsung's Proposed Combined Camera and Digital Cellular Phone

Currently only one image can be sent at a time. However, in the future the user will be able to send multiple images simultaneously using a single click. This is similar to the "Add to Basket" interfaces that are being used at web-based shopping sites. In this approach, selected items are loaded into a virtual basket or cart, and once you are done shopping you select "Check Out" to purchase all of the items simultaneously. For *Compadre*, multiple images can be selected and entered into the queue, and when the user is ready to connect to the remote server, then simply selecting the "Translate" button will connect the SmartPhone to the remote server, which in turn will process the images and return the resulting translation. The images will be transmitted back to the user using an HTML format. The users can then scroll through these images and save or delete them as is desired.

One item of note is that *Compadre's* Hybrid Translator can be configured to handle different types of input using a variety of methods. For voice-based input, the context in which words are used is readily available. This often is not the case with the camera-

based mode. For example, the words "Post Office" without context could be interpreted as a "Pole that is stuck in the ground" and "A place where people work." Thus, SpeechGear configured the translator to be dominated by a Translation Memory (TM) mode versus Machine Translation (MT). In TM, the translator uses a known set of previously translated phrases to achieve accurate outputs. Such an approach is used very often if for example an operator's manual has been previously translated, but has now been updated and thus needs to be translated once again. In the case of the camera-based system, the TM approach will be used to enter signs and information, such as the Post Office example that was stated above. Thus, SpeechGear is in the process of building the TM database to include signage typically seen on signs.

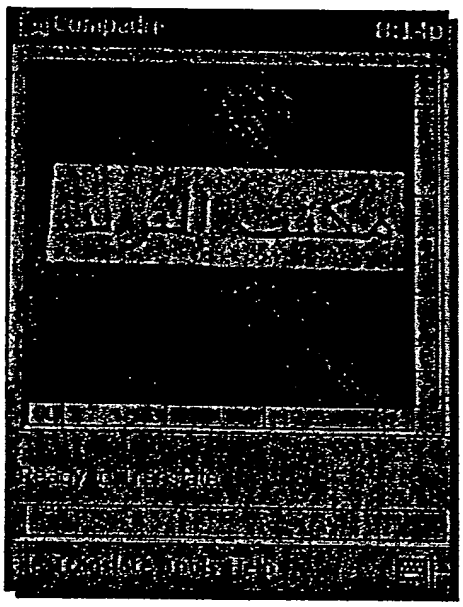


Figure 11: Example of Touchscreen Interface for Stand-Alone Mode

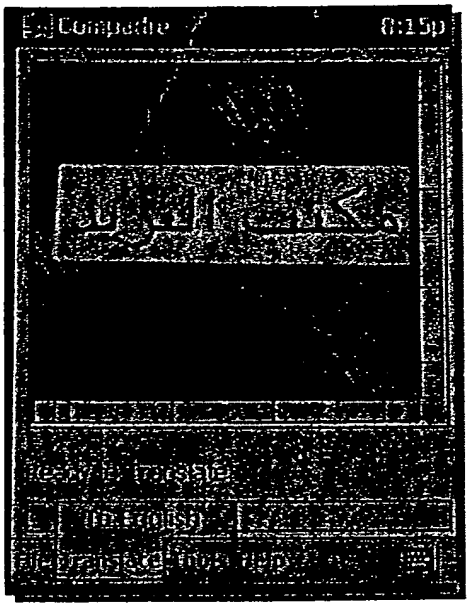


Figure 12: Example of Touchscreen Interface for Stand-Alone Mode

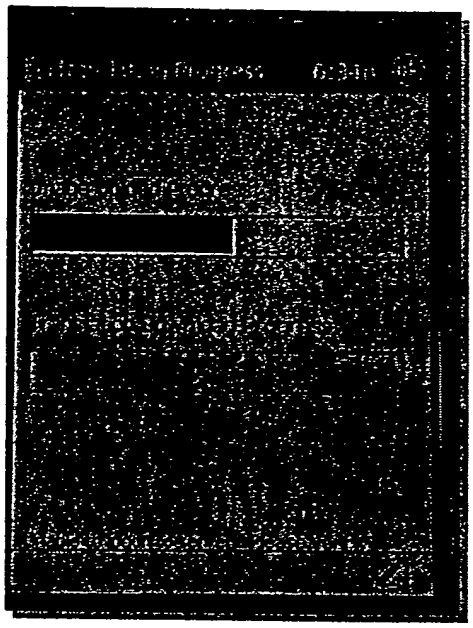


Figure13: Graphical User Interface for Viewing Results of Translation

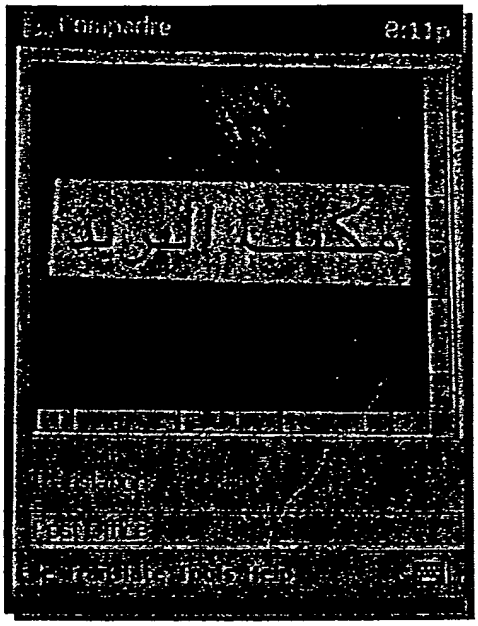


Figure14: Graphical User Interface for Viewing Results of Translation

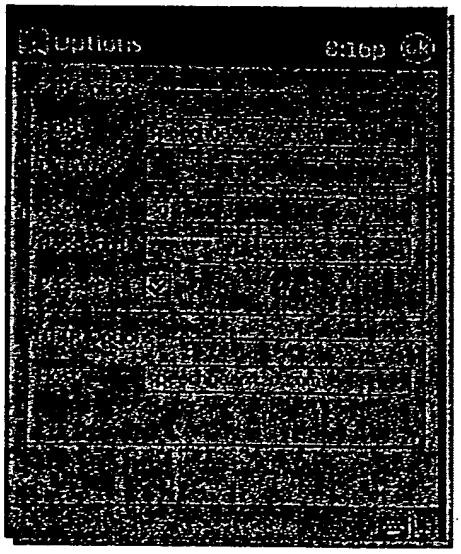


Figure15: Graphical User Interface for Viewing Results of Translation

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☑ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.